# **NEWSLETTER #5** SEPTEMBER 2023



## **INSIDE THIS ISSUE**



## PAGE 2

**Orchestrator Testbed** Targeted workflow(s)



## PAGE 3

**CINECA** Cloud Infrastructure Virtual cluster on G100



## PAGE 4

**ACROSS orchestrator** validation A first glance on experimental results

## Welcome to the **ACROSS Newsletter #5!**

The fifth edition of the ACROSS Newsletter important covers one aspect of anv development: validating its behaviour in real contexts.

As such, this issue focuses on the setup of a testbed with the purpose of validating and assessing the capabilities of the ACROSS orchestration stack. To this end, a full-fledged small test infrastructure has been set up upon **CINECA G100 Cloud-based resources.** 

The validation is going to take place during this period and will be based on application workflows derived from Pilot use cases.

Want to know more? Keep up to date on all project news @across\_project and sign up for the Newsletter so you never miss out.

Questions? Contact info@acrossproject.eu











This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955648. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Italy, France, Czech Republic, United Kingdom, Greece, Netherlands, Germany, Norway.



## **Orchestrator Testbed**

The ACROSS orchestration stack is a complex system integrating diverse software components, each with a specific purpose and providing specific features. Together, these components allow the ACROSS orchestrator to allocate HPC (and Cloud) resources, scheduling the execution of workflow's application steps on top of them, this all by addressing a more *deterministic execution* when compared to the traditional submission of jobs to the batch scheduler.

Validating the capability of this software stack is of primary importance to ensure the proper execution of workflows in production environments. As such, we are planning an extensive experimental campaign, where workflows derived from Pilot use cases and properly scaled down will be used.

### Workflow(s) target

ACROSS project aims at providing a flexible, deterministic, efficient way of executing modern workflows composed of steps belonging to diverse domains (i.e., numerical simulations, machine learning model training and inference, HPDA, etc.). To assess the capability of the orchestration stack to support their execution, we target to perform a large experimental campaign.

The first workflow being part of this experimental campaign is derived directly from the ACROSS aeronautic use case referred to as Turbine use case (WP5 – see Figure 1). In this case, the workflow is composed by numerical simulations (LES), and a high-performance data analytics (HPDA) procedure.

To reduce the timeframe for executing an experiment, the entire workflow has been down-scaled, and the URANS (high-fidelity simulations) have been removed, this all without impacting on

the complexity of the workflow in terms of step heterogeneity.

Figure 2 provides a graphical representation of the steps involved in the workflow. The last period of the project will see the number of workflows increasing, thus further demonstrating the benefit of the developed orchestration stack.

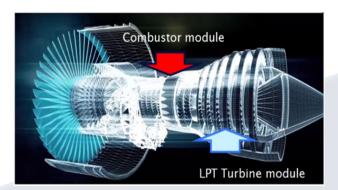


Fig. 1 – ACROSS Turbine use case

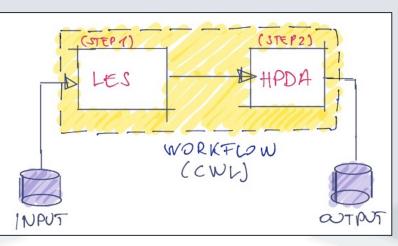


Fig. 2 - Graphical representation of the down-scaled Turbine workflow



## **CINECA Cloud Infrastructure**

Supercomputers are becoming more heterogeneous and modular in order to accommodate different users' needs (AI modelling, data analytics, etc.). As such, resource provisioning models are evolving beyond the queuing system, thus including today Cloud computing models. CINECA is following this approach, by supporting Cloud provisioning model through its ADA Cloud system.

#### **Cloud service and HPC resources**

The following points apply (see also Figure 3):

- HPC and CLOUD are crossing paths.
- HPC centres are aware of the change in the user requests and provide more "cloudy" services (<u>https://fenix-ri.eu/).</u>
- 1 Major hyperscalers (Azure, Google, etc.) are expanding their offer to provide HPC services based on high-end technology (e.g., NVIDIA GPUs) and their own HPC hardware (e.g., AWS Graviton, Google TPU).

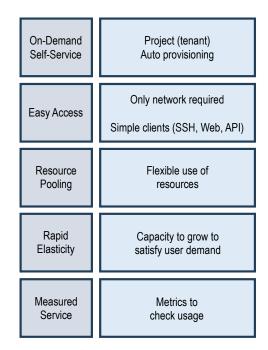


Fig. 3 – Cloud resource provisioning features

| <b>Cloud-HPC</b> | benefits |
|------------------|----------|
|------------------|----------|

| User requested features    | Motivation                      | Implementation  |
|----------------------------|---------------------------------|---|
| Agility                    | Quickly deploy IT projects      | Composable services are provided through OpenStack          |
| Reliability                | Disaster Recovery<br>Continuity | DR is not offered.<br>Physical infrastructure redundancy    |
| Scalability and Elasticity | Meet user requirements          | Constant growth in capacity                                 |
| Performance                | HPC like<br>SLAs                | Offered<br>Not offered, but HD available on<br>working days |
| Ease of maintenance        | Infrastructure managing         | Clear division of roles (users/admins)                      |
| Security and compliance    | GDPR<br>ISO27001                | Offered   |
| User Data                  | Ownership                       | Users remain owners of their data                           |



## **CINECA ADA Cloud**

The HPC cloud infrastructure, named ADA Cloud is based on OpenStack Wallaby. It provides:

- 71 interactive OpenStack nodes each equipped with 2 x CPU Intel Cascade Lake 8260, 24 cores each, clocked at 2.4 GHz, and sporting 768 GB of RAM and 2TB of SSD storage.
- 1 PB of Ceph dedicated raw storage (i.e., fully based on NVMe/SSD).

This cloud infrastructure is tightly connected both to the LUSTRE storage of 20 PB raw capacity, and to the GSS storage of 6 PB seen by all other infrastructural resources. This setup enables the use of all available HPC systems (i.e., Tier-0 Marconi, Tier-1 Galileo100), addressing HPC workloads in conjunction with cloud resources. It is worth remarking the following points:

- OpenStack Horizon dashboard: <u>https://adacloud.hpc.cineca.it</u>
- User authentication via CINECA IdP, based on Keycloak OpenID. Additional Fenix IdP is going to be available.
- User Guide: <a href="https://wiki.u-gov.it/confluence/display/SCAIUS/UG3.5%3A+ADA+Cloud+UserGuide">https://wiki.u-gov.it/confluence/display/SCAIUS/UG3.5%3A+ADA+Cloud+UserGuide</a>

Multiple cloud images are provided, but users can use their own:

 Multiple cloud flavours are provided, ranging from 1 up to 96 vCPUs, to allocate the entire compute node capability with a single VM. This is to address computing power demanding user workflows.

#### **ACROSS Virtual Cluster**

On top of the (Galileo-100) ADA Cloud, a smallscale virtual cluster has been created. The cluster sports a moderate number of (virtual) resources which are backed by high-performance physical ones. The total amount of compute cores (vCPUs) is 288, backed with a total storage volume of 3.0 TiB, and connected with a (high-performance) networking using a flatten topology. For the purposes of the validation process, these resources have been arranged as 19 VMs as follows (see also Figure 4):

- 1 VM equipped with 16 vCPUs used to export the 3.0 TiB of storage as a NFS shared volume.
- 2 VMs equipped with 8 vCPUs each hosting the SLURM batch scheduler, along with other services and orchestration modules (i.e., WARP, StreamFlow, HyperQueue).

• 16 VMs equipped with 16 vCPUs each, used as compute nodes.

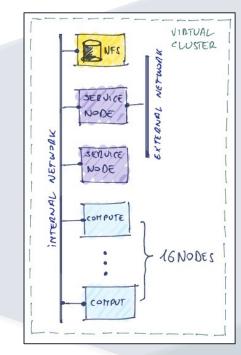


Fig. 4 – ACROSS Virtual Cluster architecture



## ACROSS orchestrator validation

An extensive campaign of experiments aimed at comparing the execution of workflows with and without the ACROSS orchestration stack has been planned.

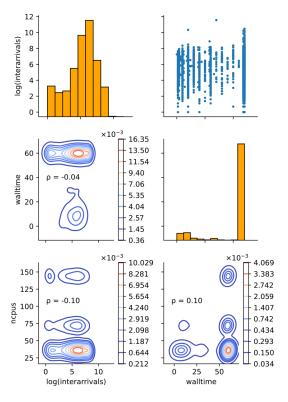


Fig. 5 – Jobs' distribution model

#### A first glance on experimental results

The workflow (it is meant to be the first one of a set of workflows derived from pilot use cases) as described on page 2 has been implemented using StreamFlow/WARP description format (CWL). A first set of runs have been performed in a clean environment (not loaded with jobs) in order to ensure the correctness of the implementation (see Figure 6). During the next months, a large experimental campaign will be performed to assess the benefit provided by the ACROSS orchestration stack to have a 'deterministic' workflow execution.

|  | <pre>step /LiS/_transfer_/run_script received inputs ['0']</pre>   |
|--|--|
|  |  |
|  |  |
|  | Step /LES/sim_file_token_transformer_received inputs ['0', '0', '0', '8']<br>COUPLETED Step /LES/cup script-taken_transformer  |
|  | COMPLETED Step /LES/In_1cfuetoken-transformer<br>COMPLETED Step /LES/in_tile-token-transformer   |
|  | Step /LES Step /LES/stm_/java_files received inputs ['0']  |
|  | Step / ESSclamaterectivel induction of the state of             |
|  | Step / ESSstate_state_states transformer received inputs [10], [0], [0]  |
|  | Detrieving available locations for job /LES/0 on slurg.vol30_LES/LES.  |
|  | Available locations for   ab /LES/0 on sturm vol00 LES/LES are ['locathost'].  |
|  | COMPLETED Step /LES/java_files-taken-transformer   |
| 2023-09-15 15:34:54.136 DEBUG                                      | Called '/reservation' MARE API with parameters workflow_id=test_hpda, step_id=/LES. Reservation id is:   |
| spSuser_104  |  |
|  | Job /LES/0 allocated on location slurm_vg100_LES/LES/Localhost   |
| /userextermal/wpSuser/tmp/67ma723                                  | EXECUTING command mkdir -p /g100_scratch/home/userexternal/wp5user/tmp/d3d3d5753-efe2-4x32-x666-13b528c3e1f8 /g100_scratch/home<br>20-5652-46cc-b5f7-d31b050f9cdc /g100_scratch/home/userexternal/wp5user/tmp/3872602d-k250-40e8-x8cb-53bx31b7fbfe 2x61 an slum_v  |
| g108_LES/LES/localhost<br>2023-09-15 15:34:54,184 DEBUG            | EXECUTING command mkpir -== /a100 scratch/hpre/sserexternal/wp5user/tma/a3d3g753-efe2-4a32-a6b6-13b628c3e1f8/8666667b-9a4a-45f5  |
| -8a85-5e72376efa08 2x61 pg slurm                                   | vg184 LES/LES/Lesthest   |
|  | DXEUTING romand mkdir -p /g100_scratck/hore/sserexternal/wpbusnr/tng/d3d3d7u3-efr2-4s32-a6b6-13b62%c3e1f8/129a2378-23ac-4c88   |
| -8954-fad940408055 2>61 on slurn                                   | vg180_LES/LES/LES/LES/LES/LES/LES/LES/LES/LES/   |
|  | EXEUUTING command mkdir -p./g100_scratch/home/userexternal/wp5user/tmg/d3d3d753-efe2-4a32-a6b6-13b628c3e1f8/97/f878f-3f73-48c5   |
| -ae8d-468aad4a8b27 2×61 om slurm_                                  |  |
|  | EXECUTING command mkoir -p /g100_scratch/home/userexternal/wp5user/tmp/63d36753-efe2-4632-6666-13b628c3e1f8/fc3082fb-6698-47ac   |
| -al3a-586093983147 2>51 om slurm_<br>2023-09-15 15:34:54.289 DEBUG | /v0100_LEx/LEx/Lex/Localnost<br>EXEUTING conmand mxcir -p /g100_scratch/home/userexternal/wp5user/tmp/c3d3o753-ete2-4a32-a6b6-13b628c3e1t8/a325catt-57e7-4bd3  |
| -8a27-e45ec6e56973 2×51 on sturn                                   |  |
| 2023-09-15 15:34:54,214 DEBUG                                      | DXEOUTING commain mxir = 0 /0100 scratch/hore/userexternal/wp5user/tmp/03d30753-ere2-4a32-a8b6-13b828c3e118/7331ab19-626a-4177   |
| -a885-15424194800f 2×51 on sturn.                                  |  |
|  | EXECUTING command mxcir -p /g100_scratch/hore/userexternal/wp5user/tmp/c3d3c753-efe2-4e32-a6b6-13b628c3ei18/007f1eb7-badf-42ec   |
| -b214-cb57a6ab95e0 2+51 on slurn                                   | vg180_LES/LES/Lest thost   |
|  | <pre>Step /LES/schedulereceived inputs ['0', '0']</pre>  |
|  | EXECUTING command in -snf /p100_scratch/wp5/CML-DEMO/LES-CML-Test.sin /p100_scratch/home/userexternal/wp5user/tmp/d3d3d753-efe   |
|  | P=23ac=4e88=89bf=fad9484a8855/LES=CHL=Test.sim 2>41 on slorm_vg100_LES/LES/locathost   |
| 2023-09-15 15:34:54.234 DEBUG                                      | EXECUTING command in -snf /g100_scratch/wp5/CWL-DEN0/extractData2HPDA.java /g100_scratch/home/userexternal/wp5user/tmp/d3d3d75   |
|  | /1078/-3173-49c5-ae8d-468aad4a8b27/extractData2HPDA.jaya 2×61 on sturm_vq100_LES/LES/localhest<br>EXECUTING command in -snf /c100_scratch/xp5/CML-0EN0/exportCS/ByTable.jaya/g100_scratch/home/userexternal/wp5user/tmp/d3d3d75  |
|  | totilina commano in −smi /gide_strator/wpb/tom=Ourv/exportsSmy/mate.java /gide_strator/name/usersternat/wpouser/imp/didato/<br>N00/fb=b00+47m==31a=5000309143/expontSSN9/fbh.java 2x43 un slumn voi0e LES/LOSIDentHest   |
|  | NW2/T0-B090-4/7E-alla-NW293/98/T4//82075.5209/1816.300 / 2001 E Surm_Vglu@_LC2/LC3/localest<br>(OPYIMS /0100 scratch/home/userstermal/AdSuser/Au5-hoda/va//srints/run LES.x4 an local file-system to /0100 scratch/home/use  |
|  | <pre>certain /gital_triangle_interventions/interventions/interventions/intervention_intervention</pre> |
|  | EXECUTING contains test = "/always cratch/hore/asservatornal/aphysics/taz/ald/a/3/a-cfr/-4a32-abb-13b5/26-3o1f8/86006/b-/b-/a4a-45f5   |
|  | Al as flum value (15/15/16/16/16/16/16/16/16/16/16/16/16/16/16/  |
| 2023-09-15 15:34:54.258 DEHUS                                      | COMPLETED Step /LES/schedule   |
| 2023-09-15 15:34:54.251 DEBUS                                      | EXELUTING command in snf /g100_scratch/wp5/CWL-DEP0/exportCSVByTable.dat /g100_scratch/hone/userexternal/wp5user/tnp/d3d3d753  |
|  | Scoff 67c7 4bd3 8b27 c45cc8c96973/cxpartCSVByTable.dat 2x61 on slurm_vg108_LE5/LE5/Lacalhost   |
|  | EXECUTING contand in -snf /g100_scratch/wp5/CML-DEMO/extractData2HPDA_v1.8.dat /g100_scratch/home/userexternal/wp3user/tmp/d3d   |
|  | 8/7331ab19-626a-4171-a886-154241948001/extractData2HPDA_v1.8.cat 2>61 on slurm_vg100_LE5/LE5/Lccalhost   |
|  | EXECUTING command in -inf /g100_scratch/wp5/CML-DEM0/sequence0fSteps.dat /g100_scratch/home/userexternal/wp5user/tmp/d303d753-   |
| erez-1032-0000-130528c3e118/00711                                  | epr-pagt=122C=210=CD574Bap958W/sequenceUF5Ceps_dat_2001_en_sturn_vc180_LE57LE57L6st Uncst  |
| erez=1032=2000=130628G3e118/00311                                  | leb7-bad1=42ec=6214=cb57a6ab95e0/sequence0f5teps.dat 2>61 on slurm_vg100_LE5/LE5/localhost   |

Fig. 6 – Execution of the small-scale Turbine-derived workflow

Visit our website <u>www.acrossproject.eu</u> to see all our latest developments

#### Loading the virtual cluster

One important aspect of performing orchestration validation through an experimental campaign concerns the creation of an 'initial load' on the virtual cluster, such that it reproduces the load condition of a real production system within a controlled environment. To this end, we used historical data collected on a real cluster to derive a statistical model of the jobs submitted. In from the available data, particular, specific features of the submitted jobs have been extracted and used to fit a multi-variated distribution of the jobs over time. This distribution is then used to randomly picking up job's features, create a job with those features and submit it to the system. Once a certain number of jobs have entered the virtual cluster, validation workflow is executed, and the system performance/behaviour is evaluated. Figure 5 provides some graphical information of the jobs' distribution model.